

J Mol Model (2010) 16:1709–1720  
DOI 10.1007/s00894-010-0806-5

## ORIGINAL PAPER

# Environment influences on the aromatic character of nucleobases and amino acids

Piotr Cysewski · Beata Szeffler

Received: 9 December 2009 / Accepted: 9 July 2010 / Published online: 29 July 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Geometric (HOMA) and magnetic (NICS) indices of aromaticity were estimated for aromatic rings of amino acids and nucleobases. Cartesian coordinates were taken directly either from PDB files deposited in public databases at the finest resolution available ( $\leq 1.5 \text{ \AA}$ ), or from structures resulting from full gradient geometry optimization in a hybrid QM/MM approach. Significant environmental effects imposing alterations of HOMA values were noted for all aromatic rings analysed. Furthermore, even extra fine resolution ( $\leq 1.0 \text{ \AA}$ ) is not sufficient for direct estimation of HOMA values based on Cartesian coordinates provided by PDB files. The values of mean bond errors seem to be much higher than the  $0.05 \text{ \AA}$  often reported for PDB files. The use of quantum chemistry geometry optimization is strongly advised; even a simple QM/MM model comprising only the aromatic substructure within the QM region and the rest of biomolecule treated classically within the MM framework proved to be a promising means of describing aromaticity inside native environments. According to the results presented, three consequences of the interaction with the environment can be observed that induce changes in structural and magnetic indices of aromaticity. First,

broad ranges of HOMA or NICS values are usually obtained for different conformations of nearest neighborhood. Next, these values and their means can differ significantly from those characterising isolated monomers. The most significant increase in aromaticities is expected for the six-membered rings of guanine, thymine and cytosine. The same trend was also noticed for all amino acids inside proteins but this effect was much smaller, reaching the highest value for the five-membered ring of tryptophan. Explicit water solutions impose similar changes on HOMA and NICS distributions. Thus, environment effects of protein, DNA and even explicit water molecules are non-negligible sources of aromaticity changes appearing in the rings of nucleobases and aromatic amino acids residues.

**Keywords** Aromaticity · Nucleobases · Amino acids · HOMA · NICS

## Introduction

Environment effects, whether physical or chemical in nature, can lead to significant structural alterations to molecules directly exposed to such forces. These changes are of particular interest for molecules important from a biochemical perspective. Nucleic acid bases and amino acids, as important building blocks of polynucleotide and polypeptide chains, respectively, are compounds of particular significance. All nucleobases and four amino acid residues possess aromatic rings that can be affected by environmental heterogeneity. These influences can be as different as explicit solvent molecules, dense packing in solids, particular molecular surroundings in polymeric neighborhoods and a variety of organic or inorganic ligands directly interacting with the solute. Straightforward contacts between molecules in closest

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-010-0806-5) contains supplementary material, which is available to authorized users.

P. Cysewski (✉) · B. Szeffler  
Department of Physical Chemistry, Collegium Medicum,  
Nicolaus Copernicus University,  
Kurpińskiego 5,  
85-950 Bydgoszcz, Poland  
e-mail: piotr.cysewski@cm.umk.pl

P. Cysewski  
Department of General Chemistry,  
University of Technology and Life Sciences in Bydgoszcz,  
Seminarnyja 3,  
85-326 Bydgoszcz, Poland

proximity, or chemical modification or indirect influence via intermolecular interactions can lead to sizable alterations of physico-chemical properties including aromaticity. This term is associated with a cyclic  $\pi$ -electron delocalization imposing an increase of stability as measured by aromatic stabilization energy [1–3], small bond length alterations within the ring skeleton [4, 5, 25],  $\pi$ -electron ring current formation after exposition to an external magnetic field [6–10], and, typically, reactivity toward substitution rather than addition reactions [11–13]. This proves that aromaticity is a multidimensional phenomenon [14] and justifies utilization of different measures of aromatic character. Most aromaticity indices are not independent, and many reports have documented correlations between different aromaticity criteria [15–20]. Despite the fact that any mutual relationships between aromaticity indices depend strongly on the selection of molecules in the probe [15, 21, 22], the correspondence between different indices can be significant. For example, in the case of porphyrins [19] the magnetic index changes monotonically with the geometry-based index. Similarly, rings in benzenoid hydrocarbons exhibit a linear correlation between HOMA and NICS [20]. In addition, energies of ring stabilization in benzenoid hydrocarbons correlate with NICS and HOMA values [23, 24]. Moreover, the magnetic index of aromaticity correlates best with geometry-based indices for five-member heterocyclic compounds [15]. Although not all aromaticity indices can be applied to heterocycles, e.g., the  $A_j$  index formulated by Jugl and Francois [25, 26], or the bond alternation coefficient (BAC) formulated by Krygowski and co-workers [27], there are many indices that can be used effectively for this particular class of compound. The HOMA index has especially proved to be universal tool that not only offers excellent agreement with chemical experience and intuition [28] but also describes electron delocalization of fragments or the  $\pi$ -electron system as a whole. An additional and very important advantage of using HOMA as the criterion of aromaticity is its potential applicability to experimental bond lengths resulting from a variety of diffraction measurements. These data are deposited in public databases providing molecular geometries of organic molecules [29] (Cambridge Structural Database), amino acids [30] (Protein Bank Database) or nucleobases [31] (Nucleic Acid Database). The usefulness of this kind of application has proven effective in numerous studies [32–38].

Intermolecular interactions are an important factor affecting aromaticity. For example, most heterocyclic and some carbocyclic compounds increase their aromaticities with an increase in medium polarity [39]. Differences in net of hydrogen bonding may also lead to serious alterations of aromaticity, as for example those that take place in the p-nitrosophenolate anion in the crystalline state [5]. Furthermore, the aromaticity of nucleobases involved in Watson-Crick pairs can also be changed significantly with respect to isolated monomers [40].

The systematic study of the effect of hydrogen bonding on the aromaticity of phenol and p-nitrophenol undertaken by Krygowski et al. [41] demonstrated a monotonous relationship between ring aromaticities and H-bond strength. The perturbation of aromaticities imposed on  $\pi$ -electron systems by alkali metal cations [42–47], alkaline earth cations [32, 48–50] as well as transition metal cation interactions with benzene [51–53] were also intensively studied. Aromaticities of nucleobases and amino acids have also been described based on different aromaticity criteria [54–65].

The aim of this study was to analyze environment influences on the aromatic character of four amino acid residues [phenylalanine (PHE), tyrosine (TYR), tryptophan (TRP) and histidine(HIS)] and four DNA bases [adenine (ADE), guanine (GUA), thymine (THY) and cytosine (CYT)] with the aid of geometric (HOMA) and magnetic (NICS) indices. The results provide evidence that the environment can significantly affect aromatic ring flexibility, as quantified by these aromaticity indices. The direct utilization of molecular geometries taken from public databases is also discussed and compared to QM/MM optimized structures of aromatic rings immersed in native environments. The separation of influences on aromaticity in actual protein or DNA surroundings from artifacts related to experimental inaccuracies is another important aspect considered in this study. To the best of our knowledge, this is the first study demonstrating the sensitivity of HOMA indices to experimental errors for the aromatic compounds considered.

## Methods

The Cartesian coordinates of amino acids were extracted from the Brookhaven Protein Database [30, 66] after selection of only those structures meeting the quality standard of finest resolution ( $\leq 1.5$  Å). Similarly, the coordinates of nucleobases were retrieved from the Nucleic Acid Database using the same accuracy criteria. Aromatic rings have been probed quite extensive and the statistical details of data distribution are provided in the electronic supplementary material (Tables S1 and S2 for the two classes of compounds analysed). Two criteria of aromaticity were applied, namely HOMA [67–69] and NICS [70] indices. These criteria were selected due to their wide usage [19, 70–75], simple meaning [68, 69] and direct applicability to both experimental [32–38, 76] and theoretical [41, 77–82] geometries. Three types of calculation were performed in this project to estimate the aromatic properties of the compounds analysed. In the first step, an estimation of HOMA sensitivity to the experimental errors in bond lengths was performed. Then HOMA and NICS(1) values were estimated directly using experimental geometries.

Finally, these two criteria of aromaticity were calculated for geometries resulting from quantum chemistry optimizations within a QM/MM hybrid approach [68, 69]. The model analysis was performed to generate HOMA distributions related to geometry distortions induced by experimental inaccuracies. The original [55, 68] formula defining HOMA was modified by introducing the following bond length changes:

$$HOMA(\omega_j) = 1 - \frac{1}{n} \sum_i \alpha (R_{opt} - (R_i + \omega_j))^2$$

where  $\omega_j$  stands for the error of  $i$ -th bond length (in case of  $\omega_j=0$ , the classical definition of HOMA is recovered),  $R_{opt}$  corresponds to reference values [69] (1.388 for C–C bond and 1.334 for C–N bond),  $R_i$  represents the optimized bond length for the isolated monomer in the gas phase, and  $\alpha$  is the normalization constant [69] equal to 257.7 and 93.52 for C–C and C–N bonds, respectively. The  $\omega_j$  values were in the range of  $<-0.1 \text{ \AA}, +0.1 \text{ \AA}>$  with steps of  $0.005 \text{ \AA}$  and were used for the alteration of bonds lengths. All possible combinations of bonds deformations were taken into account by successive changes of one, two, three, etc., up to all bonds constituting the ring. For example, in the case of six-membered rings there are 704 unique combinations for any value of bond length errors. Five-membered rings may be affected 222 ways, while for rings formed by nine centres the number of combinations increases dramatically up to 19,008. In this way, a series of HOMA values were generated corresponding to all possible distortions of the optimized structures by increasing or decreasing the bond lengths in the ring skeleton. The structures were then grouped according to the distribution of the average bond length errors, as shown by the smoothed histograms in Figs. 1, 5 and 6. In the second part of our study, full gradient optimizations were performed for representative sets of aromatic rings. The hybrid quantum chemistry/molecular mechanics approach was used for systems containing particular aromatic amino acid residues or nucleobases with initial geometries taken directly from PDB files. The geometry of each particular residue (PHE, TYR, HIS, TRP or any nucleoside) was optimized at the B3LYP/6-311+G\*\* level of theory. The rest of the system was treated classically using AMBER force-field [85, 86] embedded according to ONIOM [83, 84] methodology. In addition, explicit water solutions were modeled by immersing of nucleobases or aromatic amino acid residues within a drop of water  $20 \text{ \AA}$  in diameter. The QM part in these QM/MM studies comprised the analyzed monomer along with all water molecules within  $3 \text{ \AA}$  vicinity from any atom of the monomer. Inclusion of at least the first hydration layer within the QM region is important, as previously demonstrated for the fenol-fluoride system [81]. In the case of

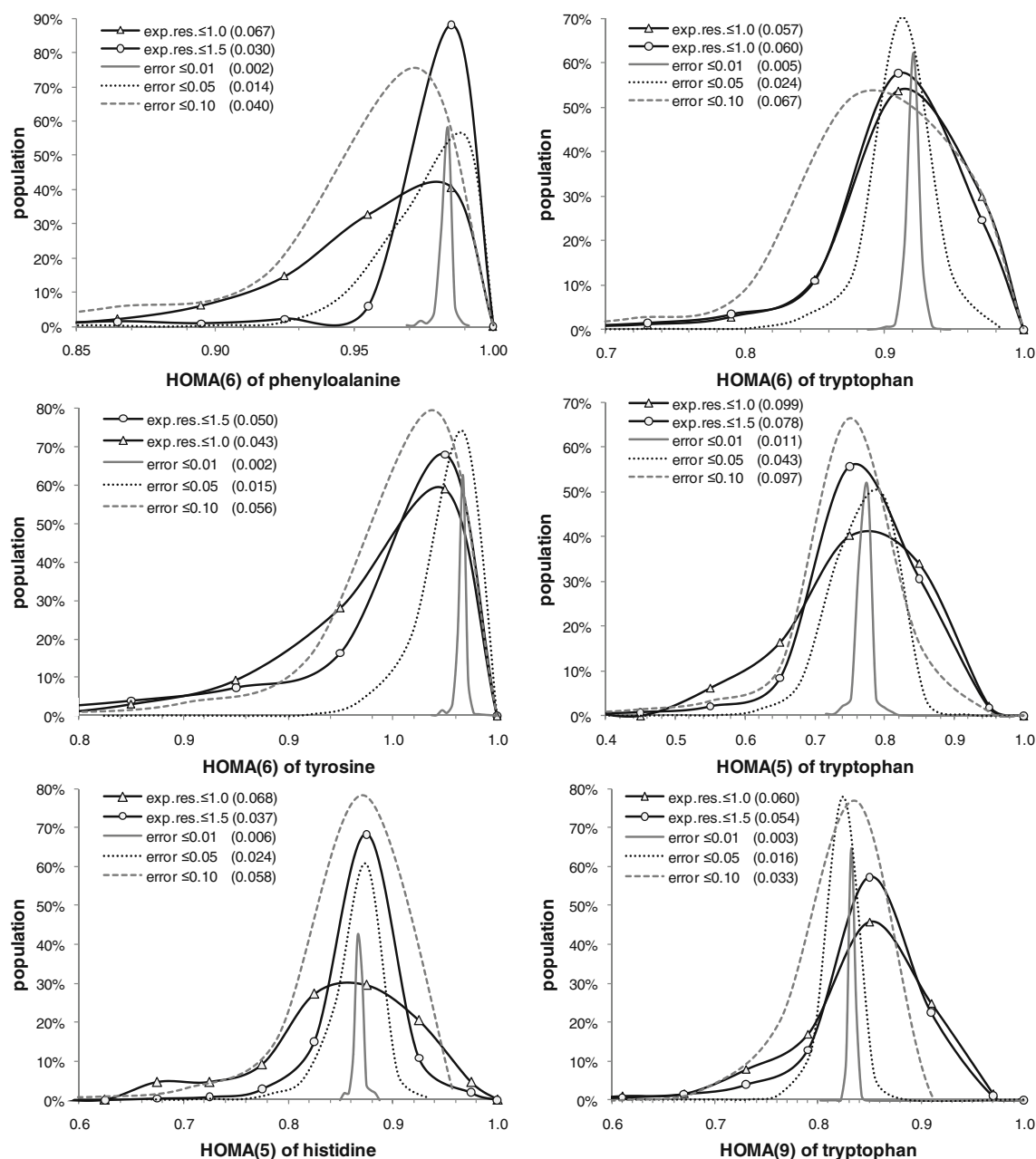
proteins or DNA, however, only monomers were included in the QM region. Nucleobase interactions via hydrogen bonds, inter- or intra-strand stacking would require a significant increase in the computation resources needed. For example, in the case of nucleobases, the QM region would increase six-fold after inclusion of only closest contacts. This poses serious technical problems due to the rapid increase in computational resources needed. Although such a simplified model can potentially be a source of error due as it can ignore important influences (mainly dispersions), it is nevertheless proposed here as a reasonable compromise between accuracy and cost of computations. Some tests were performed for amino acid molecules inside proteins, and inclusion of the nearest neighboring amino acids in the QM part only slightly affected HOMA values (increasing HOMA by  $<0.01$  unit). Extending the QM region led to a significant increase in computation costs, as well as difficulties in convergence of SCF procedures and minimization protocols. Thus, the proposed model properly serves the purposes of this paper. All quantum chemistry and hybrid QM/MM calculations were performed using the Gaussian03 package [87].

## Results and discussion

Definition of the HOMA index reveals its significant sensitivity to bond length deviations from reference values. The purpose of using the square-like relation was to allow quantification of aromaticity changes induced by even minute alterations from the ideal structure of benzene. However, if coordinates are taken directly from X-ray diffraction patterns of crystals, inaccuracies arising from experimental procedures and interpretation can also affect the geometric criterion of aromaticity. Naturally, these changes are not necessarily related to the alteration of aromatic character but can simply represent artefacts of inaccurate geometries. Only in cases where the observed variations in HOMA are higher than those related to individual bond length errors, can the environment influence on aromatic character be considered as significant. In order to visualize the influence of experimental inaccuracies on HOMA values, a series of model calculations was performed as described in [Methods](#). Since the different compounds analysed herein differ in aromatic nature and ring size, the analysis was performed separately for each class of aromatic rings.

### Experimental errors affecting HOMA values of amino acids

Figure 1 presents the distributions of HOMA values for each of the analysed aromatic rings of amino acid residues for three selected values of mean error of bond lengths. The



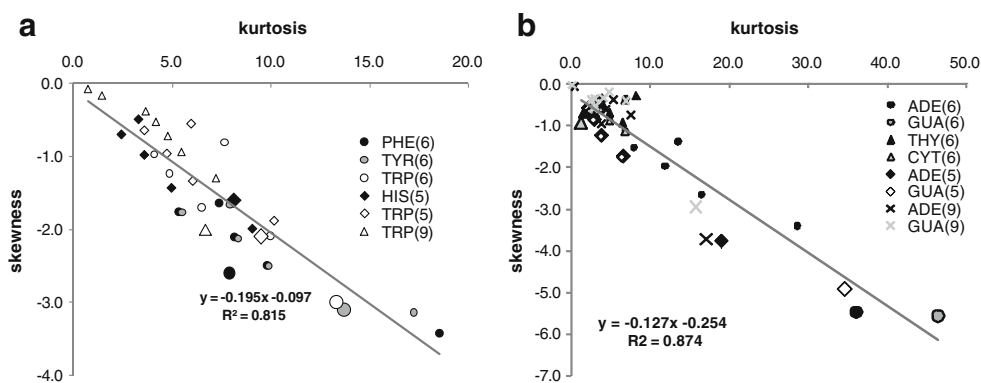
**Fig. 1** Heterogeneity of the structural index of aromaticity (HOMA) calculated for four aromatic amino acids residues. Two sets of plots correspond to the systematic analysis of mean bond length errors (gray lines) and to protein environment (black lines with open circles or triangles). Only high quality structures (resolution better than 1.5

Å) from the Protein Database [30] were used. For detailed statistics, see Table S1. Standard deviations of the analysed HOMA values distribution are given in brackets. The amount of mean bond length errors is indicated in the legends.  $HOMA(x)$  Aromaticity of the  $x$ -membered ring

significant influence on HOMA values of these inaccuracies is clearly visible, irrespective of residue type. Firstly, a considerable negative skew characterizes all distributions apart from the most precise case where the bond length inaccuracies do not exceed 0.01 Å. Geometries of such accuracy lead to only small HOMA alterations, as expressed by very small values of standard deviations, skewness and kurtosis. The increase in the mean value of

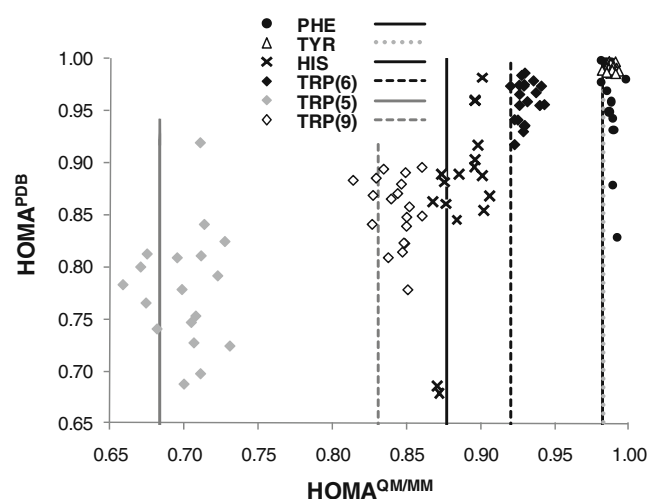
the bond length significantly affects HOMA distributions, as clearly manifested by the presence of a strong negative skew. In all cases, the highest values of the mean error of bond lengths are accompanied by the broadest distributions of HOMA values. Consequently, there is fairly linear relationship between skewness and kurtosis estimated for different bond length inaccuracies, as shown in Fig. 2. Since kurtosis and skewness are equal to zero for normal

**Fig. 2** The relationship between skewness and kurtosis for HOMA distributions of amino acid (a) and nucleobase (b) aromatic rings affected by various bond length inaccuracies. Larger symbols represent skewness and kurtosis of experimental values estimated from Cartesian coordinates taken directly from PDB files [30, 31]



distributions, these quantities can be used as an indicator of HOMA alterations observed in experimental geometries. It is evident that the aromatic rings of phenylalanine and tyrosine are most sensitive to such geometry inaccuracies since comparable bond distortions lead to the highest values of kurtosis and skewness. Tryptophan—a compound comprising two conjugated rings—is much more resistant to experimental errors of ring deformations. Aromaticity of five-membered rings is also slightly less prone to bond length errors compared to that of six-membered rings. Figure 1 also plots smoothed histograms obtained for aromatic rings of amino acid residues that are directly exposed to the influences of the protein environment. The detailed statistical characteristics are also provided in Table S1 (see electronic supplementary materials). All PDB files used for Cartesian coordinates extraction belong to the best available structures obtained with resolutions better than 1.5 Å. Despite this fact, surprisingly high values of skewness and kurtosis are observed for HOMA distributions corresponding to experimental geometries. As documented in Table S1 and Fig. 2a, all values of skewness are negative, and kurtosis can reach values of several units. Thus, either individual bond lengths differ by more than 0.1 Å or environmental effects are strong and non-negligible. It is interesting to separate these two effects in order to gain deeper insight into the observed heterogeneity of HOMA values. First of all, we found that about 2% of experimental structures were obviously incorrect. Among the several thousands of rings analysed, some were incomplete for not having all atoms in the PDB files. In other cases geometries were incorrect since their use led to non-physical values of HOMA (for example negative values for PHE). Thus, simple estimation of HOMA indices based on experimental geometries could be an alternative, direct and effective means of checking for possible errors in the Cartesian coordinates obtained via X-ray diffraction protocols. For files obtained with fine resolution ( $\leq 1.5$  Å), the estimated overall coordinate error is usually claimed to be less than 0.05 Å. However, from the plots presented in Fig. 1, it is obvious that, at least for aromatic rings, the actual

distortions must be much higher. This is also addressed in Fig. 3, where the correspondence between the HOMA values estimated using Cartesian coordinates taken directly from PDB files (denoted hereafter as  $\text{HOMA}^{\text{PDB}}$ ) and those obtained from the QM/MM optimization ( $\text{HOMA}^{\text{QM/MM}}$ ) is presented. It is evident that the broad range of  $\text{HOMA}^{\text{PDB}}$  values is significantly narrowed after QM/MM geometry optimization. In the case of tyrosine, QM/MM optimization did not introduce any changes in ring aromaticities since all  $\text{HOMA}^{\text{PDB}}$  and  $\text{HOMA}^{\text{QM/MM}}$  values differ by less than 0.012 units. The six-membered ring of tryptophan proved slightly more sensitive to optimization, with corresponding changes from 0.068 to 0.023 units. For all other amino acid residues, the geometry was changed significantly by QM/MM optimizations. For example, for phenylalanine the discrepancies in HOMA values were reduced from 0.170 to 0.017 units of HOMA. The most significant impact of QM/



**Fig. 3** Correlation between HOMA values of aromatic amino acids estimated using Cartesian coordinates taken directly from PDB files ( $\text{HOMA}^{\text{PDB}}$ ) and those obtained after QM/MM optimization ( $\text{HOMA}^{\text{QM/MM}}$ ) at ONIOM(QM:B3LYP/6-311+G\*\*//MM:AMBER) level. Vertical lines HOMA values estimated for geometries of isolated monomers optimized in the gas phase using B3LYP/6-311+G\*\* method

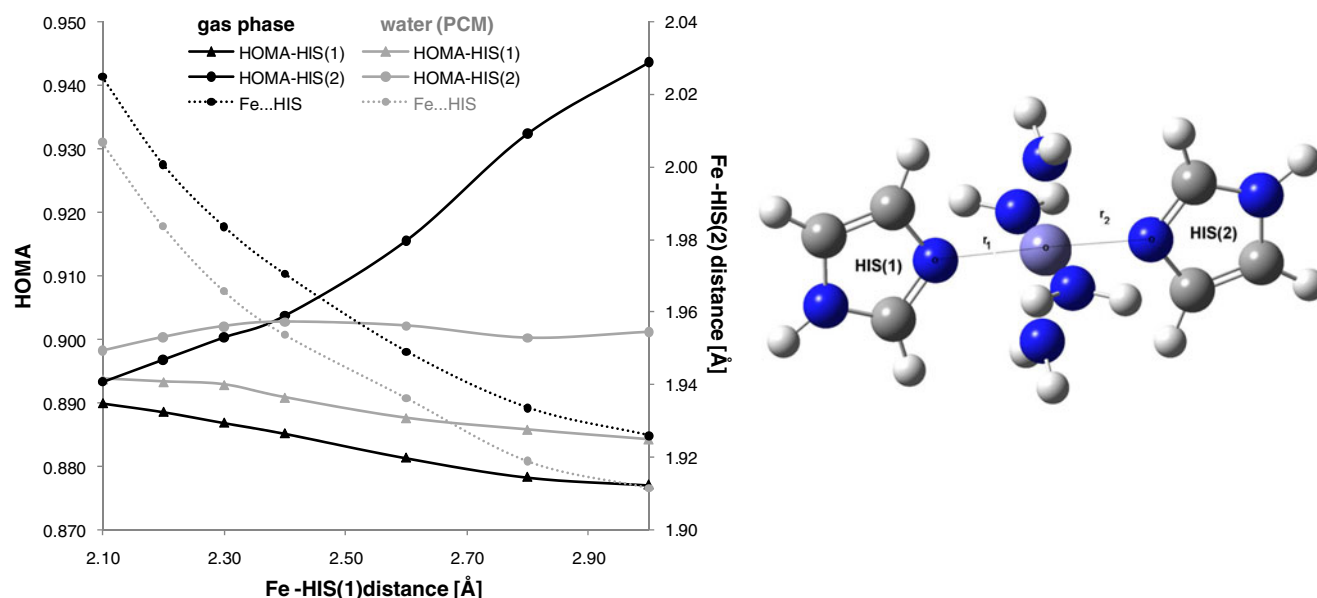


MM optimization was observed for the five-membered rings of histidine (reduced from 0.576 to 0.079) and tryptophan (decreased from 0.615 to 0.072). This suggests that most of HOMA heterogeneity coming from PDB files should not be ascribed directly to protein environment but rather to experimental inaccuracies. This is also supported by the fact that the localization of maxima in Fig. 1, indicating the most frequently occurring values of HOMA, are almost the same for PDB structures and those coming from model analysis. This conclusion is not surprising in the light of Krygowski's findings [88, 89] suggesting that  $\pi$ -electron delocalization in the aryl ring is strongly resistant to chemical modification of the substituents. However, it is known [41, 78–82] that the environment can significantly affect aromaticity, as, for example, in the case of fluoride approaching phenol in the gas phase leading to a significant reduction in the aromaticity of the phenyl ring [41, 81]. Although this effect is significantly screened [82] by the presence of discrete water molecules, it is still non-negligible, and is particularly emphasized by the broad spectrum of HOMA values obtained for different conformations of the first hydration layer around the phenol solute. Data presented in Fig. 3 suggest that protein environment affects amino acid aromaticities only slightly; however, systematic trends are observed, i.e., the HOMA values of most QM/MM structures being higher in comparison to those of the isolated monomers in the gas phase. The five-membered rings of histidine and tryptophan are characterized by the highest diversity of HOMA values obtained after QM/MM optimization. Furthermore, in the case of histidine, interesting differences in the shape of HOMA distribution plots corresponding to fine ( $\leq 1.5 \text{ \AA}$ ) and extra fine ( $\leq 1.0 \text{ \AA}$ ) resolution (see Fig. 1) are observed. One obvious reason for this fact might be the difference in the size of the two samples. However, for other amino acids this effect is almost insignificant. The second observation is that the set of highest resolution histidine residues is not quite representative since it comprises a significant portion of histidine–porphyrine complexes. Among 47 histidine residues found within PDB files corresponding to extra fine resolution ( $\leq 1.0 \text{ \AA}$ ), there are 16 rings that form complexes with a heme moiety via iron cations. These interactions might affect the aromaticity of histidine. In order to quantify the presence of iron–histidine interactions, the model system was analysed as shown in Fig. 4. It contains four amino groups and two histidine molecules forming complex with one  $\text{Fe}^{+2}$  cation. For such a simplified model of heme, the separation distance of one histidine molecule [HIS(1)] was fixed at specific values and all the remaining coordinates were optimized at the B3LYP/6-311+G\*\* level both in the gas phase and in the water solutions using the PCM model. Interestingly, the separation distance of both histidine molecules from the iron cation are not indepen-

dent and the increase in the  $r_1$  bond forces a decrease in  $r_2$  values, which suggests that HIS(2) approaches closer to the metal center. The shorter the Fe–HIS bond, the smaller the aromaticity of histidine, and this relationship is almost linear ( $R^2=0.980$ ). The polar environment modeled by continuum field of water significantly reduces the influence of the iron cation on the aromaticity of histidine. Despite the fact that Fe–HIS(2) bond lengths are slightly shorter in water solution, HOMA values are significantly less affected by this kind of environment interaction. The polar solution screens down structural changes induced by the presence of the iron cation in the model heme system. This suggests that the observed differences of HOMA distributions between extra fine and fine resolutions is only partly explained by this kind of interaction. Indeed, the average value of bond length formed between  $\text{Fe}^{2+}$  and the nitrogen atom of the histidine ring in PDB files is  $1.974 \pm 0.028 \text{ \AA}$ . However, closer inspection of PDB files leads to another interesting observation, namely there is a significant difference between the aromaticity of histidine residues located inside the protein compared to in the outer regions. The most aromatic histidine rings occupy external regions of the protein. These structures are exposed to interactions with the solution environment and can be more easily protonated, and interact more directly with the water molecules or cations that are present in the solution. Thus, these external histidine moieties are responsible for the increase of aromaticity and consequently affect HOMA distributions. On the other hand, the histidine–heme complexes are simply forced to be inside the protein volume and are prone to local interactions, which usually reduce the aromaticity of this five-membered ring. Both these effects cause the flattening of HOMA value distributions. In the case of a much larger set of structures corresponding to fine resolution ( $\leq 1.5 \text{ \AA}$ ), the population of such externally localized moieties is significantly smaller, and even reducing the resolution in PDB files to  $\leq 2.0 \text{ \AA}$  does not alter the shape of the HOMA distribution. In conclusion, it is worth emphasizing that, although it is possible to identify some environmental effects that impose alterations of HOMA of aromatic amino acid rings, the majority of observed HOMA variations are related simply to experimental inaccuracies rather than to the protein environment itself, and even extra-fine resolution ( $\leq 1.0 \text{ \AA}$ ) is not sufficient for direct assessment of aromaticities via the geometric index.

#### Experimental errors affecting HOMA values of nucleobases

Analogous investigations were performed to estimate the environmental effect on aromaticity of nucleobases. Again, model analysis was followed by extraction of the Cartesian coordinates from PDB files [30, 66] with extra fine ( $\leq 1.0 \text{ \AA}$ )

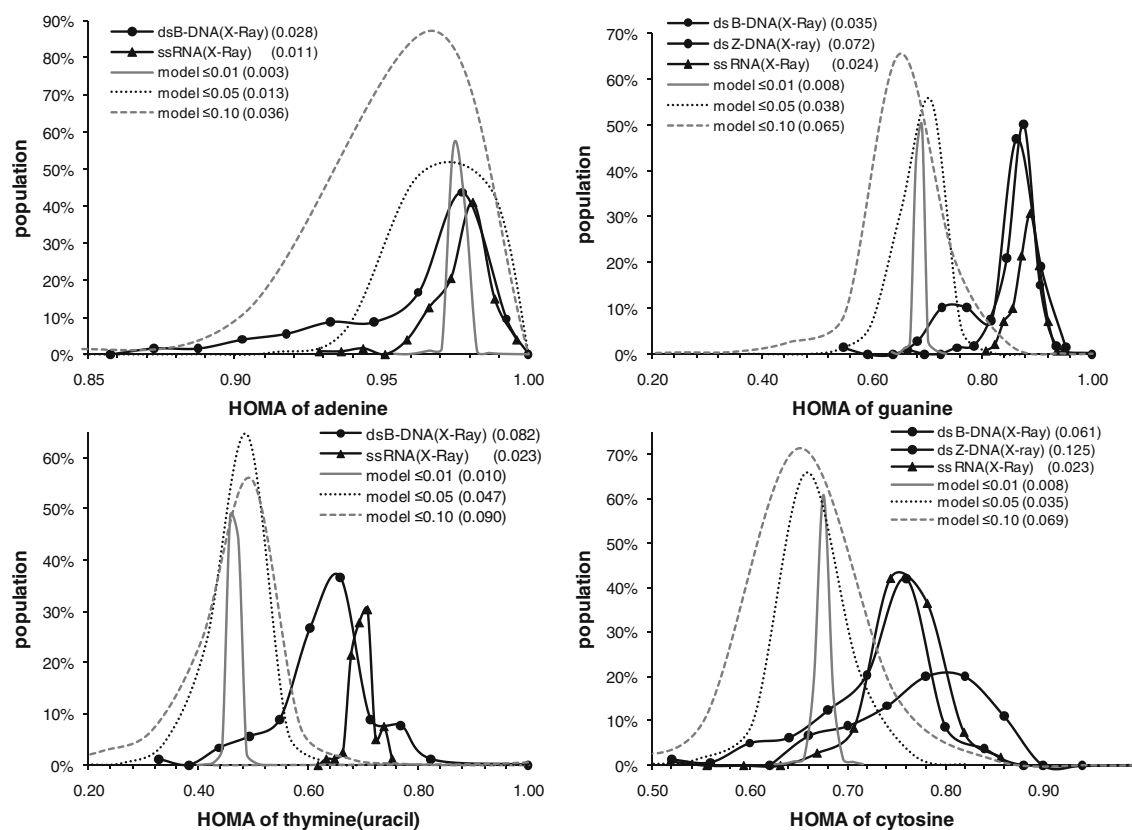


**Fig. 4** Sensitivity of histidine aromaticity to interactions with the iron cation in the model heme system. *Solid lines* HOMA values, *dotted lines* Fe-HIS distances. The distance of both histidine molecules from

$\text{Fe}^{2+}$  after full optimization is 2.041(2.026) Å, while HOMA values are 0.891(0.895). Values in brackets correspond to water solution (PCM model)

and fine ( $\leq 1.5$  Å) resolutions. As illustrated in Figs. 5 and 6, the main consequence of bond length distortions is the broadening of HOMA value distribution, without significantly affecting the positions of maxima of smoothed histograms. This observation matches that mentioned above for amino acids residues, and the linear relationships between skewness and kurtosis is also fulfilled for nucleobases (see Fig. 2b). However, the analysis of HOMA distributions from PDB files reveals a novel and interesting feature. First of all the aromaticity of pyrimidine rings in terms of the HOMA index is much higher inside the polynucleotide chain than for the isolated molecule, the only exceptions being imidazole and the purine rings of adenine. On the contrary, according to  $\text{HOMA}^{\text{PDB}}$  values, the whole molecule aromaticity of guanine is increased by 0.15 units if this nucleobase is present within a polynucleotide chain. Interestingly, the analysis of different polymorphic forms as double-stranded (ds)A-DNA, B-DNA, Z-DNA or even single-stranded (ss) RNA interiors does not alter these conclusions. The differences in HOMA value distributions in these various environments are rather minor, although some trends are visible. For example, the maxima on smoothed histograms are located at systematically higher values of HOMA for ss-RNA compared to ds-B-DNA irrespective of the nucleobase involved. Of course thymine is not present in the case of RNA and instead the uracil moiety is analyzed in Fig. 5. An even higher impact of the Z-DNA environment is observed on the aromaticities of cytosine populations. Interestingly, this effect is not observed in the case of guanine, for which all analysed

forms of polynucleotide chains induce very similar variations in HOMA distributions irrespective of the ring size. Furthermore, as was demonstrated in Fig. 2b, the skewness and kurtosis of  $\text{HOMA}^{\text{PDB}}$  value distributions are much higher for purine than for pyrimidine rings. They are even higher than for distributions obtained from model analysis. This might suggest that these rings are very sensitive to bond lengths errors or that the environment effect is significant. The analysis presented in Fig. 7 clarifies this observation. After optimization of nucleobases using the QM/MM approach, the variability of  $\text{HOMA}^{\text{QM/MM}}$  is reduced significantly if compared to  $\text{HOMA}^{\text{PDB}}$ . The most spectacular change is clearly visible in the case of thymine, for which the aromaticities of several conformations are completely incorrect if PDB coordinates are used. After geometry optimization, the structures adopt the correct geometry, clearly implicating the strange values of  $\text{HOMA}^{\text{PDB}}$  as coming from coordinates inaccuracies. Cytosine is also prone to experimental inaccuracies. After nucleoside geometry optimization, the smallest variations in  $\text{HOMA}^{\text{QM/MM}}$  values are observed for adenine (0.024) and the largest for cytosine (0.106). Furthermore, significant increases of ring aromaticities are observed inside DNA as compared to monomers in the gas phase. Only imidazole rings and the purine rings of adenine seem to behave in the opposite manner. Thus, aromaticity can be affected significantly by different contexts and structures of polynucleotide chains, leading to considerable alterations in HOMA values with respect to free monomers. The importance of the environmental effects of nucleic acid interiors on the



**Fig. 5** Heterogeneity of HOMA values characterising aromaticity of the pyrimidine rings of nucleobases. The two sets of plots correspond to systematic bond length error analysis (gray lines) and native DNA environment (coordinates directly taken from Nucleic Acid Database

[31]) (black lines with symbols). Only high quality structures (resolution better than 1.5 Å) were taken into consideration. Standard deviations of the analysed HOMA values distribution are given in brackets

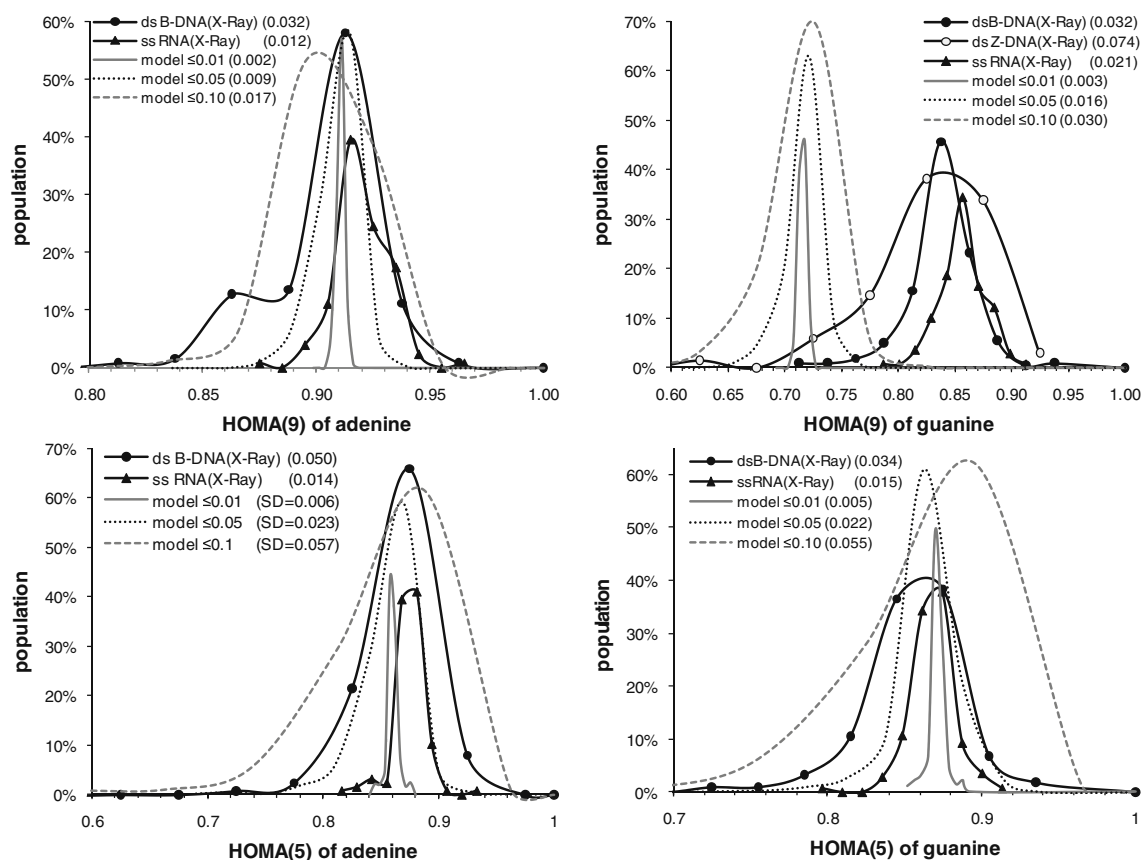
aromaticities of nucleobases is more pronounced than that of the protein environment on amino acid residues. The quite significant increase of both  $\text{HOMA}^{\text{PDB}}$  and  $\text{HOMA}^{\text{QM/MM}}$  values observed should not be ascribed merely to errors imposed on coordinates by X-ray diffraction processing, although they cannot be neglected. Furthermore, the polymorphic forms are important factors affecting the aromatic characteristics of nucleobases and the direct use of Cartesian coordinates significantly exaggerate this effect. These aspects deserve a more detailed analysis, and will be addressed in a subsequent study.

#### NICS(1) alterations imposed by environment

The final aspect analysed in this paper is the influence of the environment on aromaticity expressed by the magnetic susceptibility index. Negative values of magnetic shielding estimated 1 Å above or below the ring centre [NICS(1) or NICS(−1)] are commonly used as a measure of the magnetic criterion of aromaticity. The geometries obtained after QM/MM optimizations were used to estimate the NICS values of both amino acid residues and nucleobases. The values obtained, along with those corresponding to

isolated monomers (optimized in the gas phase) are presented in Fig. 8. Interestingly, the six-membered ring of TRP, as the most aromatic among all amino acid residues in the gas phase, is also the most affected by protein environment. Heterogeneity of aromatic character is quite high and amounts to about 8% for TRP(5) and 7% for histidine. The lowest influence of the environment is observed for the rings of phenylalanine, and tyrosine, and the six-membered rings of tryptophan. In the latter case, it reaches 0.04 of NICS units with 4% fluctuation. For most conformations comprising the aromatic rings of amino acids, a reduction of NICS values with respect to the isolated monomers was noted. The most significant influence of protein environment on NICS values was observed for the five-membered rings of tryptophan and histidine. The mean values of  $\Delta\text{NICS}$ , estimated as the difference with respect to the isolated monomers ( $\text{NICS}^{\text{protein}} - \text{NICS}^{\text{gas}}$ ) were −1.0 ppm and −0.5 ppm for TRP(5) and HIS(5), respectively. On the contrary, the average aromaticity of the six-membered ring of tryptophan is almost unaffected by protein environment ( $\Delta\text{NICS}=0.0$ ). Phenylalanine and tyrosine are characterized by  $\Delta\text{NICS}$  values of −0.4 and −0.1, respectively. All these conclusions are in good accord with



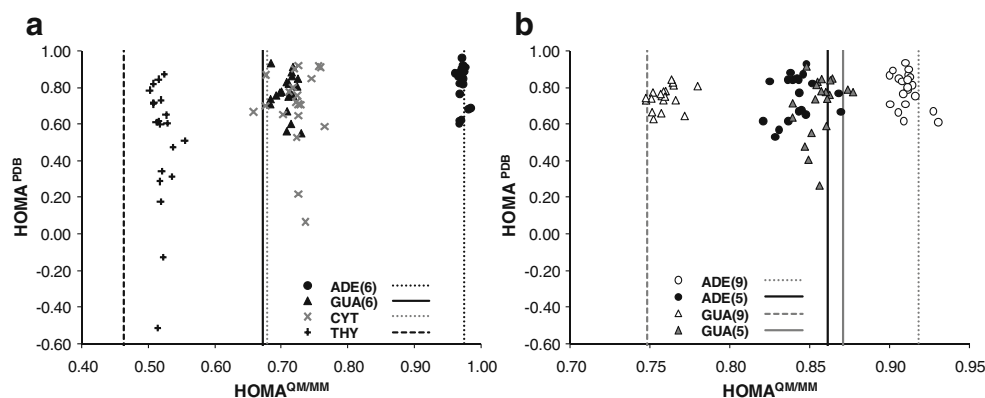


**Fig. 6** Heterogeneity of HOMA values characterising aromaticity of the purine [HOMA(9)] and imidazole (HOMA5) rings of adenine and guanine. The two sets of plots correspond to systematic bond length error analysis (gray lines without symbols) and native DNA

environment (taken from Nucleic Acid Database [31]) (black lines with symbols). Only high quality structures (resolution better than 1.5 Å) were taken into consideration. Standard deviations of the analysed HOMA values distribution are given in brackets

previously discussed observations related to the structural index of aromaticity. Thus, protein environments modestly affect the aromaticity of amino acids, although for some conformations this influence is non-negligible.

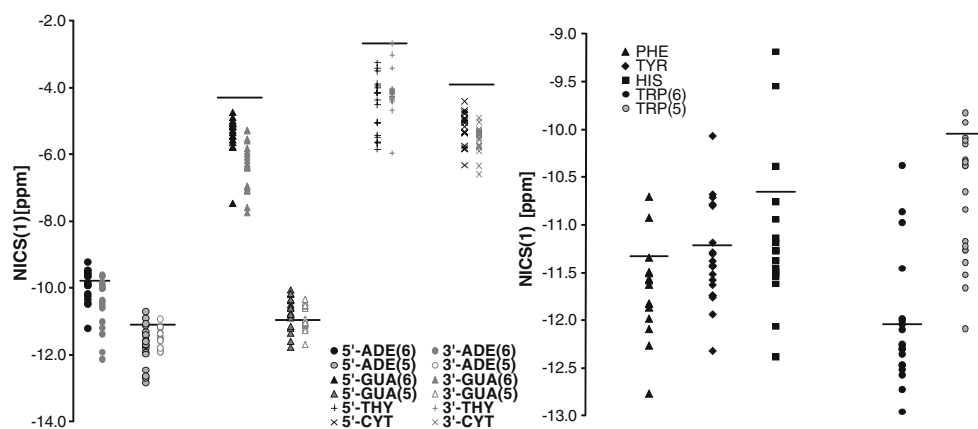
Similar results, characterising the B-DNA influence on the magnetic index of nucleobase aromaticity is presented in Fig. 8b. In the case of a double-stranded helix, two contexts, corresponding to NICS(1) and NICS(−1) distributions, are



**Fig. 7** Correlation between HOMA values of (a) pyrimidine rings and (b) imidazole/purine rings of nucleobases estimated using Cartesian coordinates taken directly from PDB files ( $\text{HOMA}^{\text{PDB}}$ ) and those obtained after QM/MM optimizations ( $\text{HOMA}^{\text{QM/MM}}$ ) at ONIOM

(QM:B3LYP/6-311+G\*\*//MM:AMBER) level. Vertical lines HOMA values estimated for geometries of isolated monomers optimized in the gas phase (using B3LYP/6-311+G\*\* level)

**Fig. 8** Heterogeneity of the magnetic measure of aromaticity [NICS(1)] of amino acids and nucleobases inside protein (PDB [30]) or DNA (NDB [31]) interiors, respectively. The lines represent reference levels of isolated monomers optimized in the gas phase

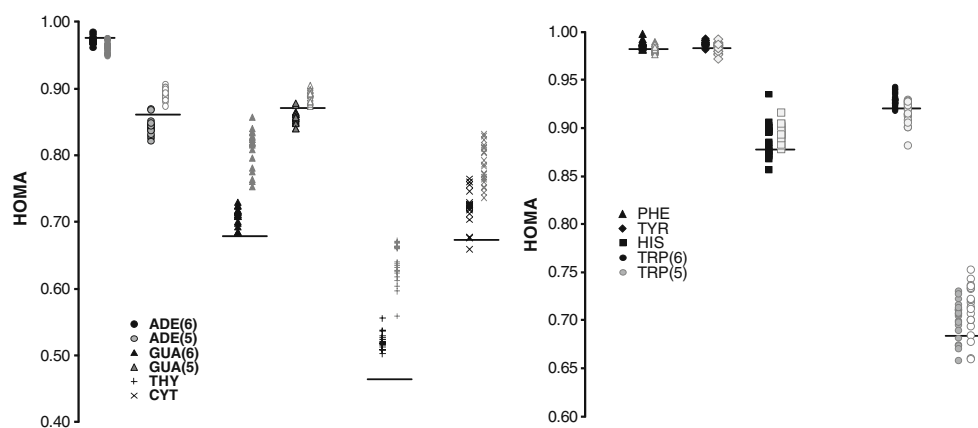


clearly identified. Here, the former is related to the 5'-side and the latter to the 3'-region of the nucleobase. Since the point located above the purine rings is above the C–C bond, the corresponding NICS values cannot be used as a measure of the whole molecule aromaticity, and only pyrimidine and imidazole rings are analysed here. According to this measure of aromaticity, the polynucleotide chain can significantly affect aromaticities of all rings, since the range of obtained values spans from the smallest variability (1.1 ppm) in the case of the 3'-imidazole ring of adenine up to the highest variety for 3'-thymine rings (reaching 3.3 ppm). Both analysed aromatic rings of adenine and the imidazole ring of guanine are characterized by positive values of mean  $\Delta$ NICS, which suggests a decrease in aromaticity with respect to the free monomer in the gas phase. For the remaining rings, a significant increase in aromaticity is expected. For example, the mean  $\Delta$ NICS values of 5'- and 3'-pyrimidine rings of guanine are equal to 1.1 ppm and 2.0 ppm, respectively. Also the remaining 5'-(3'-) pyrimidine rings are characterized by an increase in their aromatic character by  $-0.5(-0.8)$ ppm and  $-1.4(1.0)$ ppm for cytosine and thymine, respectively.

## Conclusions

The discussion presented above demonstrated that even extra fine resolution ( $\leq 1.0 \text{ \AA}$ ) is not sufficient for direct estimation of HOMA values based on Cartesian coordinates provided by PDB files obtained via X-ray diffraction of protein and DNA crystals. The values of mean bond errors seem to be much higher than the  $0.05 \text{ \AA}$  often reported for such files. This imposes serious limitations on the direct use of Cartesian coordinates, and even the best available PDB files cannot provide structures of acceptable accuracy. On the other hand, we propose the estimation of HOMA values as an effective alternative method to verify the accuracy of experimental structures. Despite the fact that environment effects are significantly overestimated if direct coordinates are taken from the PDB files, they are still non-negligible. However, the use of quantum chemistry geometry optimization is strongly advised; even a simple QM/MM model comprising only the aromatic molecule within the QM region, with the rest of the biomolecule being treated classically within the MM framework, proved a reliable method for aromaticity description inside the native

**Fig. 9** Heterogeneity of the structural index of aromaticity (HOMA) corresponding to nucleobases or amino acids inside DNA (NDB [31]), or protein (PDB [30]) interiors. Gray symbols Distributions of HOMA values estimated for rings geometries optimized in explicit water solutions using QM/MM approach, lines reference levels of isolated monomers optimized in the gas phase



environment. According to the results presented here, three consequences of environment inducing changes of structural indices of aromaticity can be observed. First of all, a broad range of HOMA values are usually obtained for different conformations. These values, and their mean, can differ significantly from those characterizing the isolated monomer. Protein and DNA interiors can also affect the values of the magnetic index of aromaticity, both by increasing the range of values related to different conformations and by altering the mean  $\Delta$ NICS values estimated with respect to isolated monomers in the gas phase. The most significant increase in aromaticity is expected for the pyrimidine rings of guanine, thymine and cytosine. The same trend is also visible for all amino acids inside proteins, but this effect is much smaller, reaching the highest value for the five-membered ring of tryptophan. Thus, environment effects are non-negligible and need further investigation to reveal the details of aromaticity changes imposed by biomolecule interiors. Interestingly, explicit water molecules also form a significantly non-homogeneous environment. Figure 9 collects values corresponding to QM/MM computations of biopolymer environments and water solutions. As discussed previously, both B-DNA and protein interiors significantly affect the aromaticity of the compounds analysed. It is also worth noting that the fluctuations of water molecules around the aromatic rings imposes flexibility on amino acid residues and nucleobase rings quantified by HOMA distributions. Although this paper is not devoted to detailed analysis of particular neighborhoods (such as B-DNA or protein sequence, context or conformation), the statistically significant probe of structures with non-zero (non-trivial) distributions of aromaticity indices provided offers serious evidence that aromaticity can be affected by environmental factors that do not induce chemical modifications of the analyzed aromatic compounds. Further investigations will be essential in order to provide a more precise description of these contributions.

**Acknowledgements** The work was supported by the Computational grant no 39 of PCSS (Poznań, Poland). The allocation of computational facilities is greatly acknowledged.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Minkin VI, Glukhovtsev MN, Simkin BYA (1994) Aromaticity and antiaromaticity electronic and structural aspects. Wiley, New York
- Cyrański MK, Krygowski TM, Katritzky AR, Schleyer PvR (2002) *J Org Chem* 67:1333–1338
- Cyrański MK, Schleyer PVR, Krygowski TM, Jao H, Hohlneicher G (2003) *Tetrahedron* 59:1657–1665
- Bird CW (1985) *Tetrahedron* 41:1409–1414
- Krygowski TM (1993) *J Chem Inf Comput Sci* 33:70–78
- Flygare WH (1974) *Chem Rev* 74:653–687
- Dauben HJ, Wilson JD, Laity JL (1968) *J Am Chem Soc* 90:811–813
- Dauben HJ, Wilson JD, Laity JL (1969) *J Am Chem Soc* 91:1991–1998
- Dauben HJ, Wilson JD, Laity JL (1971) In: Snyder JP (ed) *Nonbenzoid aromatics*, vol 2. Academic, New York, pp 167–206
- Lazzeretti P (2000) *Prog Nucl Magn Reson Spectrosc* 36:1–88
- Breslow R (1968) *Angew Chem Int Ed* 80:573–578
- Lloyd D, Marshall DR (1972) *Angew Chem Int Ed* 11:404–408
- Smith BM, March J (2001) *Advanced organic chemistry*. Wiley, New York, p 106
- Krygowski TM, Cyrański MK, Czarnocki Z, Häfeliinger G, Katritzky AR (2000) *Tetrahedron* 56:1783–1796
- Katritzky AR, Barczynski P, Musumurra G, Pisano D, Szafran M (1989) *J Am Chem Soc* 111:7–15
- Katritzky AR, Feygelman V, Musumurra G, Barczynski P, Szafran M (1990) *J Prakt Chem* 332:853–869
- Katritzky AR, Feygelman V, Musumurra G, Barczynski P, Szafran M (1990) *J Prakt Chem* 332:870–884
- Katritzky AR, Barczynski P (1990) *J Prakt Chem* 332:885–897
- Cyrański MK, Krygowski TM, Wisiorowski M, Hommes NJRVE, Schleyer PVR (1998) *Angew Chem Int Ed* 37:177–180
- Howard ST, Krygowski TM (1997) *Can J Chem* 75:1174–1181
- Katritzky AR, Karelson M, Slid S, Krygowski TM, Jug K (1998) *J Org Chem* 63:5228–5231
- Schleyer PVR, Freeman PK, Jiao H, Goldfuss B (1995) *Angew Chem Int Ed* 34:337–340
- Cyrański MK, Krygowski TM (1999) *Tetrahedron* 55:6205–6210
- Cyrański MK, Krygowski TM (1998) *Tetrahedron* 54:14919–14924
- Julg A, Francois P (1967) *Theor Chim Acta* 7:249–259
- Julg A (1971) In: Bergman ED, Pullman B (eds) *Aromaticity, pseudo-aromaticity, anti-aromaticity*, vol 3. Israel Academy of Science and Humanities, Jerusalem, p 383
- Krygowski TM, Ciesielski A, Cyrański MK (1995) *Chem Pap* 49:128–142
- Krygowski TM, Cyrański MK (2004) *Phys Chem Phys* 6:249–255
- Allen FH, Davies JE, Galloy JJ, Johnson O, Kennard O, McRae M, Mitchell EM, Mitchell GF, Smith JM, Watson DG (1991) *J Chem Inf Comput Sci* 31:187–204
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) *Nucleic Acids Res* 28:235–242
- Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B (1992) *Biophys J* 63:751–759
- Kalinowska M, Siemieniuk E, Kostro A, Lewandowski W (2006) *J Mol Struct THEOCHEM* 761:129–141
- Allen FH, Davies JE, Galloy JJ, Johnson O, Kennard O, McRae EM, Mitchell GF, Smith JM, Watson DG (1991) *J Chem Inf Comput Sci* 31:187–204
- Domenicano A, Hargittai I (1992) *Accurate molecular structures: their determination and importance*. Oxford University Press, Oxford, p 437
- Colapietro M, Domenicano A, Marciante C, Portalone GZ (1982) *Naturforschung* 37b:1309–1311
- Lister DG, Tyler JK, Hog JH, Larsen NW (1974) *J Mol Struct* 23:253–264
- Schultz G, Portalone G, Ramondo F, Domenicano A, Hargittai I (1996) *Struct Chem* 7:59–71
- Domenicano A, Schultz G, Hargittai I, Colapietro M, Portalone G, George P, Bock CW (1990) *Struct Chem* 1:107–120

39. Katritzky AR, Karelson M, Wells AP (1996) *J Org Chem* 61:1619–1623
40. Cyrański MK, Gilski M, Jaskólski M, Krygowski TM (2003) *J Org Chem* 68:8607–8613
41. Krygowski TM, Zachara JE, Szatyłowicz H (2004) *J Org Chem* 69:7038–7043
42. Sunner J, Nishizawa K, Kebarle PK (1981) *J Phys Chem* 85:1814–1820
43. Guo BC, Pumell JW, Castleman AW (1990) *Chem Phys Lett* 168:155–160
44. Wouters J (1998) *Protein Sci* 7:2471–2480
45. Amicandelo JC, Armentrou PB (2000) *J Phys Chem A* 104:11420–11432
46. Nicholas JB, Hay BP, Dixon DA (1999) *J Phys Chem A* 103:1394–1400
47. Koyanagi GK, Bohme DK (2003) *Int J Mass Spectrom* 227:563–575
48. Choi HS, Suh SB, Cho SJ, Kim KS (1998) *Proc Natl Acad Sci USA* 95:12094–12099
49. Tan XJ, Zhu WL, Ciu M, Luo XM, Gu JD, Silman I, Sussman JL, Jiang HL, Ji RY, Chen KX (2001) *Chem Phys Lett* 349:113–122
50. Rodriguez-Cruz SE, Williams ERJ (2001) *Am Soc Mass Spectrom* 12:250–257
51. Felle D, Dixon DA, Nicholas JB (2000) *J Phys Chem A* 104:11414–11419
52. Zaric SD (1999) *Chem Phys Lett* 311:77–80
53. Molina J, Dobado JA, Melchior S (2002) *J Mol Struct THEOCHEM* 589–590:337–347
54. Huang Z, Yu W, Lin Z (2006) *J Mol Struct THEOCHEM* 801:7–20
55. Compagnon I, Hagemeister FC, Antoine R, Rayane D, Broyer M, Dugourd P, Hudgins RR, Jarrold MF (2001) *J Am Chem Soc* 123:8440–8441
56. Huang ZJ, Lin ZJ (2005) *J Phys Chem* 109:2656–2659
57. Zhang ML, Huang ZJ, Lin ZJ (2005) *J Chem Phys* 122:134313
58. Hameka HF, Jensen JH (1993) *J Mol Struct THEOCHEM* 288:9–16
59. Martinez SJ III, Alfano JC, Levy DH (1993) *J Mol Spectrosc* 158:82–92
60. Samuels AC, Jensen JO, Hameka HF (1998) *J Mol Struct THEOCHEM* 454:25–30
61. Pakiari AH, Farrokhnia M, Azami SM (2008) *Chem Phys Lett* 457:211–215
62. Zhang M, Lin Z (2005) *J Mol Struct THEOCHEM* 760:159–166
63. Snoek LC, Robertson EG, Kroemer RT, Simons JP (2000) *Chem Phys Lett* 321:49–57
64. Lee KT, Sung J, Lee KJ, Kim SK, Park YD (2002) *J Chem Phys* 116:8251–8254
65. Lee Y, Jung J, Kim B, Butz P, Snoek LC, Kroemer RT, Simons JP (2004) *J Phys Chem* 108:69–73
66. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) *J Mol Biol* 112:535–542
67. Kruszewski J, Krygowski TM (1972) *Tetrahedron Lett* 36:3839–3842
68. Krygowski TM, Cyrański M (1996) *Elsevier Sci* 52:1713–1722
69. Krygowski TM, Cyrański M (1996) *Elsevier Sci* 52:10255–10264
70. Schleyer PVR, Maerker C, Dransfield A, Jiao H, Hommes NJRVE (1996) *J Am Chem Soc* 118:6317–6318
71. Schleyer PVR, Jiao H, Hommes NJRVE, Malkin VG, Malkina O (1997) *J Am Chem Soc* 119:669–676
72. Subramanian G, Schleyer PVR, Jiao H (1996) *Angew Chem Int Ed* 35:2638–2641
73. Subramanian G, Schleyer PVR, Jiao H (1997) *Organometallics* 16:2362–2369
74. Jiao H, Schleyer PVR (1996) *Angew Chem Int Ed* 35:2383–2386
75. Jiao H, Schleyer PVR, Mo Y, McAllister MA, Tidwell TT (1997) *J Am Chem Soc* 119:7075–7083
76. Brutschy B (2000) *Chem Rev* 100:3891–3920
77. Krygowski TM, Ksawery M (2001) *Chem Rev* 101:1385–1420
78. Szatyłowicz H, Krygowski TM, Panek JJ, Jezierska A (2008) *J Phys Chem A* 112:9895–9905
79. Szatyłowicz H, Krygowski TM, Hobza P (2007) *J Phys Chem A* 111:170–175
80. Krygowski TM, Szatyłowicz H, Zachara JE (2005) *J Chem Inf Model* 45:652–656
81. Cysewski P, Szeffler B, Szatyłowicz H, Krygowski TM (2009) *New J Chem* 33:1–7
82. Cysewski P, Szeffler B, Kozłowska K (2009) *J Mol Model* 15:731–738
83. Svensson M, Humbel S, Morokuma K (1996) *J Chem Phys* 105:3654–3661
84. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) *J Phys Chem* 100:19357–19363
85. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Hang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) *J Comput Chem* 16:1999–2012
86. Besler BH, Merz KM Jr, Kollman PA (1990) *J Comput Chem* 11:431–439
87. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) *Gaussian 03, Rev C.02*. Gaussian, Wallingford CT
88. Krygowski TM, Stepień BT (2004) *Pol J Chem* 78:2213–2217
89. Krygowski TM, Stepień BT, Cyrański MK, Ejsmont K (2005) *J Phys Org Chem* 18:886–891